

Supplementary Information S2. The positive dataset S_{Hyl}^+ contains 107 hydroxylysine peptide fragments and the negative dataset S_{Hyl}^- contains 836 non-hydroxylysine peptide fragments. See the main text for further explanation.

Positive dataset

Uniprot ID	Site	Sequence
Q05707	1601	PGVPGA K GERGER
P02457	1085	QGPRGD K GETGEQ
P04258	95	PGPPGM K GPAGMP
Q02388	2631	KGERGV K GACGLD
P20908	627	PGQTGP K GDRGFD
P02459	542	SGLAGP K GANGDP
Q3Y5Z3	63	PGEKGE K GDPGLV
P23805	99	VGEPGP K GDTGPR
P23805	87	VGPIGP K GDNGFV
O18495	38	VGKFYY K HKYYIK
P02459	308	AGAPGV K GESGSP
Q3Y5Z3	96	PGTPGR K GEPGES
P30754	351	SGEPGA K GPTGDT
P02745	48	RGRPGL K GEQGEP
P20908	642	AGLPGE K GHRGDP
P04258	950	RGAMGI K GHRGFP
Q9Y383	269	SHSKNP K RRSRE
P20908	846	DGPEGP K GRGGPN
Q02388	2625	MGPRGL K GERGVK
P02745	33	RAPDGK K GEAGRP
P26368	15	RQLNEN K QERDKE
P23805	63	ECPHGE K GDPGSP
P20908	774	EGPPGE K GGQGP
P23805	159	PGPSGL K GKAGAP
P12111	2322	PGYPGP K GNPGE
P20908	708	DGQPGP K GNVGPQ
P20908	897	PGANGE K GGRGTP
P02459	299	PGLDGA K GEAGAP
P02459	287	PGLPGV K GHRGYP
P02457	494	DGIAGP K GPPGER
P02459	419	DGIPGA K GSAGAP
P02457	740	RGDPGP K GADGAP
Q60994	80	AGLLGP K GETGDV
P02459	374	PGAPGA K GEAGPT
P12111	2209	RGPPGA K GNKGGP
P26368	276	LNDDQV K ELLTSF
P35248	98	AGEPGP K GERGLV

Q3Y5Z3	60	DGTPGEEKGEKGD
P12108	181	NCPPGPKGPQGLQ
Q3Y5Z3	28	EDPPLPKGACAGW
P84698	25	GYPCVNKACSGCT
O18495	27	EAGWLRKAAKSVG
P84698	18	NGWECCKGYPCVN
P02461	284	TGAPGLKGENGLP
P02459	452	TGPLGPKGQTGEP
Q00433	56	AGAAMDKSAKAPQ
P30754	546	PGVAGAKGQGGPP
P20908	795	PGPRGVKGADGIR
P02747	75	RGPKGQKGEPGLP
P20908	804	DGIRGLKGTKGEK
O18495	30	WLRKAAKSVGKIFY
Q05707	1523	PGMPGEEKGEKGD
P83415	21	IRDGVIKAAPAIE
P23805	141	KGETGPKGGVGAP
Q15848	101	PGIQGRKGEPGEG
P04258	107	PGFPGMKGHRGFD
P30943	41	VNMAMDKSAKAPV
P02745	67	TGIQGLKGDQGEP
P02457	419	SGAPGPKGNSGEP
O18495	34	AAKSVGKFYYKHK
P20908	807	RGLKGTKGEKGED
P20908	819	DGFPGFKGDMGIK
P30754	933	KGEAGGKGAKGDK
P02747	57	DGLPGPKGEPGIP
P02459	470	KGEQGPKGEPGPA
O18495	44	KHKYYIKAAWQIG
P12111	2337	NGTTGPKGIRGRR
P02459	464	PGIAGFKGEQGPK
P19999	47	DGPKGEEKGEPGQG
Q3Y5Z3	72	PGLVGPKGDTGET
Q9SE35	174	GSYSGSKGSKRRI
P30754	183	PGTPGSKGDRGQS
P02461	860	PGPQGVKGERGSP
Q4ZJN1	73	DGIRGEKGEPGAD
P02459	608	MGFPGPKGANGEP
P02745	100	RGIPGIKGTKGSP
P30754	342	MGIPGEEKGPSGEP
P23805	210	PGETGAKGESGLA
P19999	44	DGRDGPKEKGEKGP
P23805	135	PGTPGPKGETGPK
Q05707	1485	QGEPGPKGPDGPR

P02457	1097	QGDRGMK K GHRGFS
P20908	864	LGPPGEK K GKLGVP
P35248	86	RGPVGP K GENGSA
Q05707	1701	TGIKGEK K GNPGVG
Q4ZJN1	127	QGQKGD K GVEVGPT
P04258	119	DGRNGEK K GEPGAP
P30754	567	PGERGPK K GVGGSK
P02457	851	VGAPGPK K GARGSA
P02745	103	PGIKGT K GSPGNI
P20908	882	PGRQGP K GSIGFP
O18495	40	KFYKHK K YYIKAA
P20908	744	IGPPGEK K GPLGKP
P23805	162	SGLKGEK K GAPGET
P30754	939	KGAKGD K GWTGLP
P12111	2212	PGAKGNK K GGPGQP
Q05707	1698	RGLTGIK K GEKGNP
P23805	198	RGPPGLK K GDRGDP
P20908	810	KGTKGEK K GEDGFP
Q05707	1476	PGLRGP K GQQGEP
Q9Y383	266	RSRSHS K NPKRSR
P01170	120	CKNFYW K GFTSC@
P02459	620	PGKAGEK K GLPGAP
P12111	2103	RGFPGEK K GEVGEI
P02457	254	AGLPGMK K GHRGFS
P04142	47	IRDGIVK K AGPAIE
P02457	1022	DGAAGPK K GDRGET

Negative dataset

Uniprot ID	Site	Sequence
Q3Y5Z3	129	VPIRFTK K IFYNQQ
Q3Y5Z3	144	YDGTGK K FLCNIP
Q3Y5Z3	164	HITVYLK K DVKVSL
Q3Y5Z3	167	VYLKDVK K VSLYKN
Q3Y5Z3	172	VKVSLYK K NDKALL
Q3Y5Z3	175	SLYKNDK K ALLFTH
Q3Y5Z3	187	HDQFQDK K NVDQAS
Q3Y5Z3	202	VLLYLEK K GDQVWL
Q15848	33	VLLPLPK K ACTGW
Q15848	172	VYMKDVK K VSLFKK
Q15848	177	VKVSLFK K KDKAML
Q15848	178	KVSLFKK K DKAMLF
Q15848	180	SLFKKD K AMLFTY
Q60994	36	ALVPPP K GTCAGW
Q60994	195	YDQYQEK K NVDQAS

Q4ZJN1	49	DGRDGA K GDKGDA
Q4ZJN1	52	DGAKGD K GDAGEP
Q4ZJN1	66	HPGGPG K DGIRGE
Q4ZJN1	85	DGRVEA K GIKGDP
Q4ZJN1	88	VEAKGI K GDPGSR
Q4ZJN1	99	SRGSPG K HGPKGS
Q4ZJN1	103	PGKHGP K GSIGPT
Q4ZJN1	124	TGPQGQ K GDKGEV
Q4ZJN1	148	TGPLGP K GLPGPM
Q4ZJN1	159	PMGPIG K PGPRGE
Q4ZJN1	184	RGMRGW K GDRGEK
Q4ZJN1	190	KGDRGE K GKVGEA
Q4ZJN1	192	DRGEKG K VGEAPL
Q4ZJN1	201	EAPLVP K SAFTVG
Q4ZJN1	213	GLTVIS K FPPDA
Q4ZJN1	222	PPDAPI K FDKILY
Q4ZJN1	225	APIKFD K ILYNEL
Q4ZJN1	240	YNVATG K FTCHVA
Q4ZJN1	268	VQVSLV K NGVKVL
Q4ZJN1	272	LVKNGV K VLHTKD
Q4ZJN1	277	VKVLHT K DSYMSS
Q4ZJN1	296	GIVQEL K LGDEVW
P30754	50	PMGPIG K RGPPGP
P30754	62	PAGIAG K SGDDGR
P30754	96	PGMPGP K GHRGFR
P30754	108	RGLSGS K GEGQKS
P30754	113	SKGEGQ K SGNQGP
P30754	146	ERGRDG K SGLPGL
P30754	192	RGQSGI K GAQGLQ
P30754	261	PGQAGA K GDGGPT
P30754	279	PGAPGV K GSSGPP
P30754	299	HAGEAG K RGSPGS
P30754	383	SDGRPG K DGRPGI
P30754	392	RPGIRG K DGKQGE
P30754	395	IRKDG K QGEQGP
P30754	425	ARGE P KNGAPGE
P30754	443	EQGDAG K DGETGA
P30754	488	TPGEAG K TGERGA
P30754	573	KGVGGS K GSRGDI
P30754	584	DIGPRG K AGERGK
P30754	590	KAGERG K DGERGE
P30754	612	SGLAAS K GERGDM
P30754	657	AGPAGT K GDIGFP
P30754	674	TRGATG K QGARGP

P30754	686	PRGLAGKRGLRGA
P30754	734	PAGTAGKQGVKGA
P30754	738	AGKQGVKGARGSP
P30754	749	SPGLVGKQDRGS
P30754	765	PGRDGTKGERGED
P30754	810	MGGQGMKGDGGPP
P30754	851	QEGAPGKDGLPGL
P30754	927	AGIMGMKGEAGGK
P30754	936	AGGKGAKGDKGWT
P30754	986	SQGPPGKDGQPGP
P02457	43	GLTYNDKDVKPE
P02457	47	NDKDVWKPEPCQI
P02457	107	AGVEGPKGDTGPR
P02457	160	SYGYDEKSAGVAV
P02457	217	PAGPPKNGDDGE
P02457	226	DDGEAGKPGRPGQ
P02457	266	SGLDGAKGQPGPA
P02457	275	PGPAGPKGEPGSP
P02457	341	PGAAGAKGETGPQ
P02457	386	DGQPGA KGATGAP
P02457	431	PGAPGNKGD TGAK
P02457	437	KGDTGAKGEPGPA
P02457	457	PAGEEGKRGARGE
P02457	509	PGAVGPKGSPGEA
P02457	527	AGLPGA KGLTGSP
P02457	541	SPGPDGKTGPPGP
P02457	575	MGFPGPKGAAGEP
P02457	583	AAGEPGKPGERGA
P02457	601	AVGAAGKDGEAGA
P02457	646	PPGEAGKPGEQGV
P02457	698	PGNDGAKGDAGAP
P02457	731	AGLPGA KGDRGDP
P02457	748	ADGAPGKDGLRGL
P02457	770	AGAPGDKGEAGPP
P02457	815	DGQPGA KGETGDA
P02457	824	TGDAGAKGDAGPP
P02457	892	PPGPAGKZGSKGP
P02457	896	AGKZGSKGPRGET
P02457	923	PGPPGEKGS PGAD
P02457	973	PSGEPGKQGPSGA
P02457	1051	PVGPAGKNGDRGE
P02457	1141	SAGAAGKDGLNGL
P02457	1197	PQPPQEKAHDGGR
P02457	1225	EVDTTLKSLSQOI

P02457	1242	SPEGTRKNPARTC
P02457	1252	RTCRDLKMCHGDW
P02457	1259	MCHGDWKSGEYWI
P02457	1277	CNLDAIKVYCNME
P02457	1299	QATIAQKNWYLSK
P02457	1305	KNWYLSKNPKEKK
P02457	1308	YLSKNPKEKKHVW
P02457	1310	SKNPKEKKHVWFG
P02457	1311	KNPKEKKHVWFGE
P02457	1360	NVTYHCKNSVAYM
P02457	1374	HDTGNLKKALLLQ
P02457	1375	DTGNLKKALLLQG
P02457	1413	HTGAWGKTVIEYK
P02457	1419	KTVIEYKTTKTSR
P02457	1422	IEYKTTKTSRLPI
P02459	30	HGQDVQKAGSCVQ
P02459	73	IICEDMKDCLSPE
P02459	103	SGQPGPKGQKGEF
P02459	106	PGPKGQKGEFGDI
P02459	113	GEPGDIKDIVGPK
P02459	119	KDIVGPKGPPGPQ
P02459	140	RGDRGDKGEKGAP
P02459	143	RGDKGEKGAPGPR
P02459	190	AGGFDEKAGGAQM
P02459	259	DDGEAGKPGKSGE
P02459	262	EAGKPGKSGERGP
P02459	574	DAGPQKVGPSGA
P02459	616	ANGEPGKAGEKGL
P02459	634	LRGLPGKDGETGA
P02459	731	PGTDGPKGAAGPA
P02459	764	AGIAGPKGDRGDV
P02459	773	RGDVGEKGPEGAP
P02459	781	PEGAPKDGGRGL
P02459	803	AGANGEKGEVGPP
P02459	857	QGEAGQKGDAGAP
P02459	884	TGVTGPKGARGAQ
P02459	925	PPGPSGKDGPKGA
P02459	929	SGKDGPKGARGDS
P02459	956	AGPPGEKGEFGDD
P02459	1055	DGAAGVKGDRGET
P02459	1084	PAGPIGKQGDRGE
P02459	1118	QGPRGDKGETGEA
P02459	1130	AGERGLKGHRGFT
P02459	1174	PVGPSGKDGANGI

P02459	1231	GLGQREKGPDPLO
P02459	1260	EVDATLKSLNNQI
P02459	1287	RTCRDLKLCHPEW
P02459	1294	LCHPEWKSGDYWI
P02459	1312	CTLDAMKVFCNME
P02459	1333	NPASVPKKNWSS
P02459	1334	PASVPKKNWSSK
P02459	1340	KNWSSKSKDKKH
P02459	1342	WWSSKSKDKKHIW
P02459	1344	SSKSKDKKHIWFG
P02459	1345	SKSKDKKHIWFG
P02459	1394	NITYHCKNSIAYL
P02459	1408	EAAGNLKALLIQ
P02459	1409	AAGNLKALLIQG
P02459	1435	FTYTVLKDGTCKH
P02459	1440	LKDGCTKHTGKWG
P02459	1444	CTKHTGKWGKTM
P02459	1447	HTGKWGKTMIEYR
P02459	1456	IEYRSQKTSRLPI
P04258	8	YEAYDVKSGVAGG
P04258	73	AIGPSGKDGESGR
P04258	128	PGAPGLKGENGVP
P04258	194	PGSPGAKGEVGA
P04258	239	DGSPGGKGEVGA
P04258	280	AAGEPGKNGAKGD
P04258	284	PGKNGAKGDPGPR
P04258	305	PGIAGPKGEDGKD
P04258	310	PKGEDGKDGSPGE
P04258	347	NGLPGEKGPPGDR
P04258	394	GPGSNGKPGPPGS
P04258	428	MGFPGPKNNDGAP
P04258	436	NDGAPGKNGERGG
P04258	454	PQGPAGKNGETGP
P04258	473	TGSPGDKGDTGPP
P04258	499	PPGENGKPGEPGP
P04258	506	PGEPGPKEAGAP
P04258	518	PGIPGGKGDGAP
P04258	551	PGPEGGKGAAGPP
P04258	584	PGGPGPKGDKGEP
P04258	587	PGPKGDKGEPGSS
P04258	601	VDGAPGKDGPRGP
P04258	623	AGQPGDKGESGAP
P04258	668	NGEPGAKGERGAP
P04258	677	RGAPGEKGEGGPP

P04258	704	PGPQGVKGERGSP
P04258	745	SSGAPGKDGPFGP
P04258	767	PGISGPKGDSGPP
P04258	821	PGPQGIKGENGKP
P04258	826	IKGENGKPGPSGQ
P04258	875	DGAPGAKGDRGEN
P04258	994	PSGPPGKDASGH
P12108	51	DGIDGDKGSPGAP
P12108	63	PGSPGAKGEPGAP
P12108	77	PDGPPGKPGLDGL
P12108	87	DGLTGAKGSRGPW
P12108	99	WGGQGLKGQPGLP
P12108	137	IGVPGPKGDPGPD
P12108	154	PPGPPGKPGPPGH
P12108	190	QGLQGLKGHRGRP
P12108	207	EPGQQGKQGPKGD
P12108	211	QGKQGPKGDVGV
P12108	241	PGMAGPKGETGPA
P12108	250	TGPAGYKGMVGTI
P12108	268	PGREGPKGPPGDP
P12108	277	PGDPGEKGELGGR
P12108	295	QGDIGPKGDMGLP
P12108	306	LPGIDGKDGTGPI
P12108	316	PGIPGVKGTAGQP
P12108	343	PGQPGSKGGPGDK
P12108	349	KGGPGDKGEVGAR
P12108	385	AGVPGLKGDRGER
P12108	406	AGQSGPKGEQGP
P12108	424	QGLPGVKGDKGSP
P12108	427	PGVKGDKGS PGKT
P12108	432	DKGSPGKTGPKGS
P12108	436	PGKTGPKGSTGDP
P12108	451	HGLAGVKGEKGES
P12108	454	AGVKGEKGESGEP
P12108	463	SGEPGPKGQQGIQ
P12108	529	IIDVVLKMMQEQL
P12108	543	EVAVSAKRAALGG
P12108	595	IGNIGPKGKRGEK
P12108	597	NIGPKGKRGEKGE
P12108	601	KGKRGEKGERGDT
P12108	634	GHALAGKDGERGP
Q05707	7	MKIFQRKMRYWLL
Q05707	51	SIQISWKAPRGKF
Q05707	56	WKAPRGKFGGYKL

Q05707	61	GKFGGYKLLVTPT
Q05707	71	TPTSGGKTNQLNL
Q05707	83	LQNTATKAI IQGL
Q05707	104	QIIAYNKDKESKP
Q05707	106	IAYNKDKESKPAQ
Q05707	109	NKDKESKPAQQQF
Q05707	118	QQQFRIKDLEKRR
Q05707	122	RIKDLEKRRKDPKP
Q05707	124	KDLEKRRKDPKPRV
Q05707	127	EKRKDPKPRVKVV
Q05707	131	DPKPRVKVVDRGN
Q05707	148	SSPEEVKFVCQTP
Q05707	194	FDVGSEKTRIGLA
Q05707	218	LNAFSTKDEVIEA
Q05707	231	VRNLPHYKGGNTLT
Q05707	250	IFENSFKPEAGSR
Q05707	261	SRTGVSKIGILIT
Q05707	270	ILITDGKSQDDII
Q05707	295	LFAIGVKNADVNE
Q05707	346	EQDREIKASAHAI
Q05707	382	APGNVEKYRVVYY
Q05707	394	YPTRGGKPDEVVV
Q05707	410	VSSTVLKNLMSLT
Q05707	462	ENSMRVKWDVAVPG
Q05707	488	GLAGDEKEMKIGE
Q05707	491	GDEKEMKIGETHT
Q05707	591	ITTFPLKGLTPLT
Q05707	655	ADEGLHKLWIPV
Q05707	665	IPVYGGKTEEVVL
Q05707	672	TEEVVLKEEQDSH
Q05707	754	TSSLRVKWDISDS
Q05707	796	QNNLLLKPLLPDT
Q05707	805	LPDTEYKVTVTPI
Q05707	825	SVSAPGKTLPSSG
Q05707	857	PPSSPVKGYRIVY
Q05707	864	GYRIVYKPVSVPG
Q05707	898	GMDYNVKIFASQA
Q05707	916	ALTGMVKTFLFLGV
Q05707	928	VTNLQAKHVEMTS
Q05707	960	SLQDRQKQESTVG
Q05707	984	QPDSEYKISVYTK
Q05707	990	KISVYTKLQEIEG
Q05707	1004	SVSIMEKTQSLPT
Q05707	1023	PTIPPAKEVCKAA

Q05707	1027	PAKEVCKAAKADL
Q05707	1030	EVCKAAKADLVFM
Q05707	1050	GDEFNFKIISFLY
Q05707	1064	TVGALNKGITDGT
Q05707	1086	DPRTEFKLNAYKT
Q05707	1091	FKLNAYKTKETLL
Q05707	1093	LNAYKTKETLLDA
Q05707	1101	TLLDAIKHISYKG
Q05707	1106	IKHISYKGGNTKT
Q05707	1111	YKGGNTKTGKAIK
Q05707	1114	GNTKTGKAIKYVR
Q05707	1117	KTGKAIKYVRDTL
Q05707	1136	TRRGIPKVIVVIT
Q05707	1152	SQDDVNKISREMQ
Q05707	1183	LVSIGSKPSARHV
Q05707	1199	DDFDAFKKIEDEL
Q05707	1200	DFDAFKKIEDELI
Q05707	1222	TCPVVHKDGIDLA
Q05707	1231	IDLAGFKMEMFG
Q05707	1241	MFGLVEKDFSSVE
Q05707	1266	PCYQLHKDALVSQ
Q05707	1311	LWEILNKNSDPLV
Q05707	1326	ILDNGGKTLTYFN
Q05707	1351	EGPEIRKIFYGSF
Q05707	1359	FYGSFHKLHIVVS
Q05707	1370	VSETLVKVVIDCK
Q05707	1376	KVVIDCKQVGEKA
Q05707	1381	CKQVGEKAMNASA
Q05707	1399	GVEVLGKMVRSRG
Q05707	1431	SWANTDKCCELPG
Q05707	1558	QRGLPGKDGSSGP
P23805	29	MTTFSQKILANAC
P23805	122	REGPSGKQGSMP
P23805	222	AEVNALKQRVTIL
P23805	243	NAFSQYKKAVLFP
P23805	244	AFSQYKKAVLFPD
P23805	257	GQAVGEKIFKTAG
P23805	260	VGEKIFKTAGAVK
P23805	266	KTAGAVKSYSDAE
P23805	279	QLCREAKGQLASP
P23805	303	MVRAQEKNAYLMS
P23805	355	EIFPDGKWNDVPC
P23805	363	NDVPCSKQLLVIC
Q9Y383	29	TTRQRIKFSDDRV

Q9Y383	37	SDDRVCKSHLLNC
Q9Y383	61	DLGECLKVHDLAL
Q9Y383	76	DYEIASKEQDFFF
Q9Y383	105	RRTEVAKKRLAET
Q9Y383	106	RTEVAKKRLAETQ
Q9Y383	122	SAEVAAKAERVHE
Q9Y383	135	LNEEIGKLLAKVE
Q9Y383	139	IGKLLAKVEQLGA
Q9Y383	154	NVEESQKVMDEVE
Q9Y383	161	VMDEVEKARAKKR
Q9Y383	165	VEKARAKKREAE
Q9Y383	166	EKARAKKREAEV
Q9Y383	186	SSFQQQKLRVCEV
Q9Y383	213	ADHFGGKLHLGFI
Q9Y383	224	FIEIREKLEELKR
Q9Y383	229	EKLEELKRVVAEK
Q9Y383	235	KRVVAEKQEKRNQ
Q9Y383	238	VAEKQEKRNQERL
Q9Y383	245	RNQERLKRREERE
Q9Y383	257	EREEREKLRRSRS
Q9Y383	289	SMSRERKRRTRSK
Q9Y383	295	KRRTRSKSREKRH
Q9Y383	299	RSKSREKRHRHS
Q9Y383	332	RSRERSKRRSSKE
Q9Y383	337	SKRRSSKERFRDQ
Q9Y383	365	PRDRDRKDKKRSY
Q9Y383	367	DRDRKDKKRSYES
Q9Y383	368	RDRKDKKRSYESA
P30943	44	AMDKSAKAPVITI
P30943	61	GCSRAPKEYTGAK
P30943	67	KEYTGAKAGGKDD
P30943	71	GAKAGGKDDEMMV
P30943	78	DDEMMVKAQSVKI
P30943	83	VKAQSVKIEVSTG
Q00433	82	KESTGGKAGGQDD
Q00433	93	DDEMMVKVASTKV
Q00433	98	VKVASTKVTVSES
Q00433	108	SESDAAKKLQEFI
Q00433	109	ESDAAKKLQEFIT
Q00433	118	EFITFEKGIDGPF
P01170	61	LSQEWSKRAVEEL
P01170	111	LPPRERKAGCKNF
P01170	115	ERKAGCKNFYWKG
P26368	20	NKQERDKENRHRK

P26368	26	KENRHRKRSHSRS
P26368	38	SRSRDRKRRSRSR
P26368	63	DRRRRSKPLTRGA
P26368	70	PLTRGAKEEHGGL
P26368	84	RSPRHEKKKKVRK
P26368	85	SPRHEKKKKVRKY
P26368	86	PRHEKKKVRKYW
P26368	87	RHEKKKVRKYWD
P26368	90	KKKKVRKYWDVPP
P26368	108	ITPMQYKAMQAAG
P26368	195	VQINQDKNFAFLE
P26368	225	FQGQSLKIRRPHD
P26368	260	VPDSAHKLFIGGL
P26368	286	TSFGPLKAFNLVK
P26368	292	KAFNLVKDSATGL
P26368	300	SATGLSKGYAFCE
P26368	328	GMQLGDKLLVQR
P26368	329	MQLGDKLLVQRA
P26368	340	RASVGAKNATLVS
P26368	408	VRDECSKYGLVKS
P26368	413	SKYGLVKSIEIPR
P26368	431	EVPGCGKIFVEFT
P26368	444	SVFDCQKAMQGLT
P26368	453	QGLTGRKFANRVV
P26368	462	NRVVVTKYCDPDS
P02745	32	CRAPDGKKGEAGR
P02745	81	PSGNPGKVGYPGP
P02745	110	GSPGNIKDQPRPA
P02745	195	FCDTTNKGLFQVV
P02745	219	DQVWVEKDPKKGH
P02745	222	WVEKDPKKGHIYQ
P02745	223	VEKDPKKGHIYQG
P02747	14	LPHLGLKLLLLLL
P02747	47	LPGAPGKDGYDGL
P02747	72	PGIRGPKGQKGEP
P02747	86	LPGHPGKNGPMGP
P02747	115	GEEGRYKQKFQSV
P02747	117	EGRYKQKFQSVFT
P02747	157	YDTSTGKFTCKVP
P02747	161	TGKFTCKVPGLYY
P02747	188	LYRSGVKVVTFCG
P02747	198	FCGHTSKTNQVNS
P04142	5	@@MNFAKILSFVF
P04142	29	APEPRWKIFKKIE

P04142	32	PRWKIFKKIEKMG
P04142	33	RWKIFKKIEKMGR
P04142	36	IFKKIEKMGRNIR
P04142	59	EVLGSAKAIGK@@
P83414	6	@KWKVFKKIEKVG
P83414	7	KWKVFKKIEKVGR
P02453	54	HDRDVWKPVPCQI
P02453	79	VICDELKDCPNAK
P02453	85	KDCPNAKVPTDEC
P02453	170	SYGYDEKSTGISV
P02453	429	SGPPGPKGNSGEP
P02453	504	DGVAGPKGPAGER
P02453	750	RGDAGPKGADGAP
P02453	861	VGAPGPKGARGSA
P02453	1107	QGDRGIKGHRGFS
P02453	1309	QPSVAQKNWYISK
P02452	67	CVCDNGKVLCDV
P02452	79	VICDETKNCPGAE
P02452	170	SYGYDEKSTGGIS
P02452	1316	KNWYISKNPDKDR
P02452	1319	YISKNPDKRHW
P02454	45	PNGETWKPVDCLI
P02454	67	CDGVLCKEDLDCP
P02454	77	DCPNPQKREGECC
P02454	583	TAGEPGKAGERGV
P02454	824	PGDTGVKGDAGPP
P02454	892	PPGPVGKEGGKGP
P02454	896	VGKEGGKGRGET
P02454	1374	QQTGNLKKSLLLQ
P02461	7	MMSFVQKGSWLLL
P02461	108	QGPQPKGDPGPP
P02461	161	YDSYDVKSGVAVG
P02461	229	PSGPAGKDGESGR
P02461	251	PGPPGIKGPAGIP
P02461	275	DGRNGEKGETGAP
P02461	461	PGVPGAKEGDKD
P02461	503	NGIPGEKGPAGER
P02461	923	PGVSGPKGDAGQP
P02461	932	AGQPGEKGSPPAQ
P02461	982	VKGESGKPGANGL
P02461	1212	AGIGGEKAGGFAP
P02461	1228	DEPMDFKINTDEI
P02461	1239	EIMTSLKSVNGQI
P02461	1256	SPDGSRKNPARNC

P02461	1266	RNCRDLKFCHPEL
P02461	1273	FCHPELKSGEYWV
P02461	1286	DPNQGCKLDAIKV
P02461	1313	PLNVPRKHWWTDS
P02461	1323	TDSSAEKKHVWFG
P02461	1324	DSSAEKKHVWFGE
P02461	1387	QASGNVKKALKLM
P02461	1388	ASGNVKKALKLMG
P02461	1391	NVKKALKLMGSNE
P02461	1401	SNEGEFKAEGNSK
P02461	1407	KAEGNSKFTYTVL
P02461	1419	LEDGCTKHTGEWS
P02461	1426	HTGEWSKTVFEYR
P02461	1435	FEYRTRKAVRLPI
P20908	8	DVHTRWKARSALR
P20908	44	QPADLLKVLDFHN
P20908	57	LPDGITKTTGFCA
P20908	69	ATRRSSKGPDVAY
P20908	79	VAYRVTKDAQLSA
P20908	88	QLSAPTQQLYPAS
P20908	107	SILTTVKAKKGSQ
P20908	109	LTTVKAKKGSQAF
P20908	110	TTVKAKKGSQAF
P20908	146	YEDHTGKPGPEDY
P20908	164	INLSDGKWHRIAL
P20908	174	IALSVHKKNVTLI
P20908	175	ALSVHKKNVTLIL
P20908	184	TLILDCKKKTTKF
P20908	185	LILDCKKKTTKFL
P20908	186	ILDCKKKTTKFLD
P20908	189	CKKKTTFKFLDRSD
P20908	288	DPEDLGKEPTPSK
P20908	294	KEPTPSKKPVEAA
P20908	295	EPTPSKKPVEAAK
P20908	301	KPVEAAKETTEVP
P20908	329	TSEGAGKEEDVGI
P20908	452	GGPRGEEKGQKGEF
P20908	455	RGEKGQKGEPAII
P20908	535	GGDAGSKGPMVSA
P20908	582	PGSGGLKGEPGDV
P20908	605	PPGPAGKPGRRGR
P20908	687	RGLLGPKGPPGPP
P20908	749	EKGPLGKPGLPGM
P20908	767	PPGHPGKEGPPGE

P20908	825	KGDMGIK ^K GDRGEI
P20908	866	PPGEK ^K GLGVPGL
P20908	905	GRGTPG ^K PGPRGQ
P20908	929	PRGITG ^K PGPKGN
P20908	933	TGKPGP ^K GNSSGD
P20908	963	TGFPGP ^K GPPGPP
P20908	989	ETGFQ ^K GTGPPGP
P20908	1034	LPGLAG ^K EGTKGD
P20908	1038	AGKEGT ^K GDPGPA
P20908	1049	PAGLP ^K GDGPPGL
P20908	1074	VGALGL ^K GNEGPP
P20908	1119	PGPAGE ^K GAPGEK
P20908	1125	KGAPGE ^K GPPQGA
P20908	1158	PGEDGDKGEIGEP
P20908	1167	IGEPGQ ^K GSKGDK
P20908	1170	PGQKGS ^K GDKGEQ
P20908	1173	KGSKGD ^K GEGQPP
P20908	1212	QGLFGQ ^K GDEGPR
P20908	1239	PGPPGE ^K GETGDV
P20908	1305	GGPPGP ^K GERGEK
P20908	1311	KGERGE ^K GESGPS
P20908	1335	PGDDGP ^K GSPGPV
P20908	1365	DGPPGD ^K GDDGEP
P20908	1407	EGRQGE ^K GAKGEA
P20908	1410	QGEKGA ^K GGEAGLE
P20908	1421	LEGPPG ^K TGPPIGP
P20908	1433	PQGAPG ^K PGPDGL
P20908	1473	PGLPGL ^K GDSGPK
P20908	1479	KGDSGP ^K GKEGHP
P20908	1482	SGPKGE ^K GHPGLI
P20908	1500	PGEQGE ^K GDRGLP
P20908	1515	QGSSGP ^K GEGQIT
P20908	1542	PGPPGP ^K GAKGSS
P20908	1545	PGPKGAK ^K GSSGPT
P20908	1554	SGPTGP ^K GGEAGHP
P20908	1619	GSLNSL ^K LEIEQM
P20908	1626	LEIEQM ^K RPLGTQ
P20908	1640	NPARTC ^K DLQLCH
P20908	1668	CSRDSF ^K VYCNFT
P20908	1685	TCVFPD ^K KSEGAR
P20908	1686	CVFPDK ^K KSEGARI
P20908	1697	RITSWP ^K ENPGSW
P20908	1708	SWFSEF ^K RGKLLS
P20908	1711	SEFKRG ^K LLSYVD

P20908	1762	ATGSYDKALRFLG
P20908	1792	VDGCATKKGYQKT
P20908	1793	DGCATKKGYQKTV
P20908	1797	TKKGYQKTVLEID
P20908	1806	LEIDTPKVEQVPI
P20908	1826	FGEASQKFGFEVG
P12111	33	QQQADVKNAAAD
P12111	67	FLYDVVKSLAVGE
P12111	99	LNTYRTKQEVLSH
P12111	120	GTNQTGKGLEIYM
P12111	132	MQSHLTKAAGSRA
P12111	155	LTDGHSKDGLALP
P12111	166	LPSAELKSADVNV
P12111	186	ADEGALKEIASEP
P12111	233	GDTETLKDITAQD
P12111	271	LVNLLKLPIGTQ
P12111	302	LDTYSTKAQVLGA
P12111	310	QVLGAVKALGFAG
P12111	370	YGVVALKQASVFS
P12111	414	FGDLQEKLLPYIV
P12111	430	QRHIVLKPPTIVT
P12111	443	QVIEVNKRDIVFL
P12111	470	IRDFIAKVIQRLE
P12111	505	FNTHPTKREVITA
P12111	514	VITAVRKMKPLDG
P12111	516	TAVRKMKPLDGSA
P12111	551	AAEGIPKLLVLIT
P12111	560	VLITGGKSLDEIS
P12111	573	QPAQELKRSSIMA
P12111	585	AFAIGNKGADQAE
P12111	637	PEVHSNKRDIIFL
P12111	652	GSANVGKTNFPYV
P12111	699	LNTYQTKSDILGH
P12111	783	GASQANKAELEQI
P12111	835	LAQPESKRDILFL
P12111	861	VRDFLYKIIDELN
P12111	869	IDELNVKPEGTRI
P12111	885	QYSDDVKVESRFD
P12111	896	FDEHQSKEILNL
P12111	904	EILNLVKRMKIKT
P12111	907	NLVKRMKIKTGKA
P12111	909	VKRMKIKTGKALN
P12111	912	MKIKTGKALNLGY
P12111	930	QRYIFVKSAGSRI

P12111	964	GPASNLKQSGVVP
P12111	976	PFIFQAKNADPAE
P12111	1001	AAESLPKIGDLHP
P12111	1014	QIVNLLKSVHNGA
P12111	1028	APVSGEKDVVFL
P12111	1048	SGFPLLKEFVQRV
P12111	1088	LNSYMNKQDVVNA
P12111	1156	NPSVVVKRGGAVP
P12111	1231	SPGVGGKRDVVFL
P12111	1281	QFSDDPKVEFLLN
P12111	1292	LNAHSSKDEVQNA
P12111	1305	VQRLRPKGGRQIN
P12111	1325	VSRNIFKRPLGSR
P12111	1347	VLISSGKSDDEV
P12111	1360	DPAVELKQFGVAP
P12111	1380	DQEELVKISLSPE
P12111	1403	LPSLEQKLLTPIT
P12111	1418	TSEQIQKLLASTR
P12111	1472	LNIGPSKVRVGVV
P12111	1491	FPEFYLKTYSQA
P12111	1518	SPLNTGKALEFVA
P12111	1530	ARNLFVKSAGSRI
P12111	1551	VLVLGGKSQDDVS
P12111	1635	PPSRPEKKKADIV
P12111	1636	PSRPEKKKADIVF
P12111	1637	SRPEKKKADIVFL
P12111	1694	TDEFFLKDFSTKR
P12111	1699	LKDFSTKRQIIDA
P12111	1708	IIDAINKVYKGG
P12111	1712	INKVVYKGGRHAN
P12111	1720	GRHANTKVGLEHL
P12111	1754	FVITGGKSVEDAQ
P12111	1772	LTQRGVKVFAVGV
P12111	1788	DSEEVGKIASNSA
P12111	1833	GVTDAAKACNLDV
P12111	1856	NVFVAQKGFESKV
P12111	1861	QKGFESKVDAILN
P12111	1913	QPEMLEKFRNMRS
P12111	1931	LTEDTLKVYLNKF
P12111	1936	LKVYLNKFRQSSP
P12111	1946	SSPDSVKVVIHFT
P12111	2027	LDNIAEKACCGVP
P12111	2035	CCGVPCKCSGQRG
P12111	2052	IGSIGPKGIPGED

P12111	2094	PGQRGVKGSRGFP
P12111	2121	DGEDGDKGLPGSS
P12111	2130	PGSSGEKGNPGRR
P12111	2139	PGRRGDKGPRGEK
P12111	2170	SQERGPKGETGDL
P12111	2196	GPGETGKNGGFGR
P12111	2272	TGPLGRKGEPGEP
P12111	2281	PGEPGPKGGIGNR
P12111	2309	SEGRRGKKGERGF
P12111	2310	EGRRGKKGERGFP
P12111	2355	PGIVGQKGDPGYP
P12111	2367	PGPAGPKGNRGDS
P12111	2384	ALIQSIKDKCPC
P12111	2386	IQSIKDKCPCCYG
P12111	2464	IRFADSKRKSPLL
P12111	2466	FADSKRKSPLLDK
P12111	2472	KSVLLDKIKNLQV
P12111	2474	VLLDKIKNLQVAL
P12111	2483	QVALTSKQQSLET
P12111	2500	VARNTFKRVRNGF
P12111	2510	NGFLMRKVAVFFS
P12111	2532	LREAVLKLSDAGI
P12111	2639	FQFNEMKKYIAYL
P12111	2640	QFNEMKKYIAYLV
P12111	2656	DMSDPKASQHFA
P12111	2684	ASMPPVKVEFSLT
P12111	2695	LTDYGSKEKLVDF
P12111	2697	DYGSKEKLVDFLS
P12111	2737	PNPRDLKIVVLM
P12111	2763	RVILQAKCKGYFF
P12111	2765	ILQAKCKGYFFVV
P12111	2777	VLGIGRKNVIKEV
P12111	2781	GRKVNIKEVYTFA
P12111	2796	PNDVFFKLVDKST
P12111	2800	FFKLVDKSTELNE
P12111	2834	LSPDIRKQCDWFQ
P12111	2846	QGDQPTKNLVKFG
P12111	2850	PTKNLVKFGHKQV
P12111	2854	LVKFGHKQVNVPN
P12111	2875	NPVTTTTKPVTTTT
P12111	2888	PVTTTTTKPVTTTT
P12111	2895	PVTTTTTKPVTIIN
P12111	2906	INQPSVKPAAAKP
P12111	2911	VKPAAAKPAPAKP

P12111	2916	AKPAPAKPVA AKP
P12111	2921	AKPVAAKPVATKM
P12111	2926	AKPVATKMATVRP
P12111	2937	RPPVAVKPATAAK
P12111	2943	KPATAAKPVA AKP
P12111	2948	AKPVAAKPAAVRP
P12111	2961	PAAAAAKPVATKP
P12111	2966	AKPVATKPEVPRP
P12111	2976	PRPQAAKPAATKP
P12111	2981	AKPAATKPATTKP
P12111	2986	TKPATTKPMVKMS
P12111	2990	TTKPMVKMSREVQ
P12111	3006	ITENSAKLHWERA
P12111	3035	DQSLVLKQNLTVT
P12111	3077	HGSFSTKKSQPPP
P12111	3078	GSFSTKKSQPPP
P12111	3113	TETDICKLPKDEG
P12111	3116	DICKLPKDEGTCR
P12111	3127	CRDFILKWYYDPN
P12111	3135	YYDPNTKSCARFW
P12111	3151	CGGNENKFGSQKE
P12111	3156	NKFGSQKECEKVC
P12111	3160	SQKECEKVCAPVL
P12111	3168	CAPVLA KPGVISV
Q02388	113	IRELSYKGGNTRT
Q02388	142	ARPGVPKVCILIT
Q02388	151	ILITDGKSQDLVD
Q02388	164	TAAQRLKGQGVKL
Q02388	169	LKGQGVKLFVAVGI
Q02388	176	LFAVGIKNADPEE
Q02388	184	ADPEELKRVASQP
Q02388	263	GPVTGYKVQYTPL
Q02388	459	GLEPPQKVVLPSD
Q02388	727	SAHGPEKSQLVSG
Q02388	1253	CPVYCPKGQKGEP
Q02388	1256	YCPKGQKGEPGEM
Q02388	1295	PGSATAKGERGFP
Q02388	1325	PGAPGLKGSPGLP
Q02388	1346	RGPRGPKGEPGAP
Q02388	1367	PGLPGRKGDPGPS
Q02388	1399	LPGTAMKGDKGDR
Q02388	1402	TAMKGDKGDRGER
Q02388	1443	PVGPPGKKGEKGD
Q02388	1444	VGPPGKKGEKGD

Q02388	1447	PGKKGEKGDSEEDG
Q02388	1479	PGAIGPKGDRGFP
Q02388	1494	LGEAGEKGERGPP
Q02388	1518	AGRPGA KGPEGPP
Q02388	1533	TGRQGEKGEPGRP
Q02388	1553	PAVAGPKGEKGDV
Q02388	1556	AGPKGEKGDVGPA
Q02388	1588	PGDPGPKGDPGDR
Q02388	1615	SGPPGEKGD PGRP
Q02388	1639	DGEVGEKGDEGPP
Q02388	1653	DPGLPGKAGERGL
Q02388	1672	RGPVGEKGDQGDP
Q02388	1693	PGSSGPKGDRGEP
Q02388	1718	GPGAREKGEPGDR
Q02388	1733	EGPRGPKGDPGLP
Q02388	1769	RGPAGEK GDRGPP
Q02388	1786	RSGLDGKPGAAGP
Q02388	1801	PNGAAGKAGDPGR
Q02388	1831	PPGLPGKPGEDGK
Q02388	1837	KPGEDGK PGLNGK
Q02388	1843	KPGLNGKNGEPGD
Q02388	1856	PGEDGRKGEKGDS
Q02388	1859	DGRKGEKGD SGAS
Q02388	1874	EGRDGP KGERGAP
Q02388	1912	PGGTGPKGDRGET
Q02388	1921	RGETGSKGEQGLP
Q02388	1952	LETAGIKASALRE
Q02388	1981	ERRRGPKGDSGEQ
Q02388	1992	EQGPPGKEGP IGF
Q02388	2005	PGERGLKGDRGDP
Q02388	2038	LAGEPGKPGI PGL
Q02388	2066	RGERGEKGERGEQ
Q02388	2093	PGPPGPKVSVDEP
Q02388	2113	QGPPGLK GAKGEP
Q02388	2116	PGLKGAKGEPGSN
Q02388	2128	NGDQGPKGDRGVP
Q02388	2137	RGVPGIKGDRGEP
Q02388	2166	MAGPEGK PGLQGP
Q02388	2206	QGPSGLKGE PGET
Q02388	2255	QVGETGKPGAPGR
Q02388	2267	RDGASGKD GDRGS
Q02388	2289	PGPVGPKGEPGPT
Q02388	2309	VGLPGA KGEKGAP
Q02388	2312	PGA KGEKGAPGGL

Q02388	2329	VGEPGAKGDRGLP
Q02388	2341	PGPRGEGEAGRA
Q02388	2359	PGEDGQKGAPGPK
Q02388	2365	KGAPGPKGFKGDP
Q02388	2368	PGPKGFKGDPGVG
Q02388	2388	PGPPGVKGDLGLP
Q02388	2457	PGASGLKGDKGDP
Q02388	2460	SGLKGDKGDPGVG
Q02388	2507	RGERGEGKDVGSA
Q02388	2516	VGSAGLKGDKGDS
Q02388	2519	AGLKGDKGDSAVI
Q02388	2535	PGPRGAKGDMGER
Q02388	2550	RGLDGDKGPRGDN
Q02388	2562	NGDPGDKGSKGEP
Q02388	2565	PGDKGSKGEPGDK
Q02388	2571	KGEPGDKGSAGLP
Q02388	2603	DPGSPGKDGVPGI
Q02388	2613	PGIRGEGKDVGF
Q02388	2640	CGLDGEKGDKGEA
Q02388	2643	DGEKGDKEAGPP
Q02388	2658	PGLAGHKGEMGEP
Q02388	2675	QSGAPGKEGLIGP
Q02388	2682	EGLIGPKGDRGFD
Q02388	2694	DGQPGPKGDQGEK
Q02388	2700	KGDQGEKGERGTP
Q02388	2742	EGLQGQKGERGPP
Q02388	2780	AGPRGEGEAALT
P19999	28	TCEETLKTCSVIA
P19999	63	LQGPPGKLGPPGS
P19999	79	PGSQGPKGQKGDR
P19999	82	QGPKGQKGDRGDS
P19999	94	SRAIEVKLANMEA
P19999	106	AEINTLKSLELT
P19999	108	INTLKSLELTNK
P19999	114	KLELTNKLHAFSM
P19999	122	HAFSMGKKSCKKF
P19999	123	AFSMGKKSCKKF
P19999	126	MGKKSCKKFFVTN
P19999	127	GKKSCKKFFVTNH
P19999	140	ERMPFSKVKALCS
P19999	142	MPFSKVKALCSEL
P19999	162	RNAEENKAIQEVA
P19999	169	AIQEVAKTS AFLG
P19999	199	LTYSNWKDEPND

P19999	200	TYSNWKKDEPN DH
P35248	23	DLGAEMK TLSQRS
P35248	118	ISGPAGK EGP SGK
P35248	124	KEGPSGK QGNIGP
P35248	133	NIGPQ GK PGPKGE
P35248	137	QGKPGPK GEAGPK
P35248	143	KGEAGPK GEVGAP
P35248	158	QGSAGAK GPAGPK
P35248	203	RGPPGLK GDRGAP
P35248	215	PGDRGIK GESGLP
P35248	236	MEALNGK LQRLEA
P35248	248	AAFSRYK KAALFP
P35248	262	GQSVGDK IFRAAN
P35248	277	EPFEDAK EMCRQA
P35248	308	LVTAHSK AAFLSM
P35248	322	DVGTEGK FTYPTG
P35248	362	NGQWNDK ACGEQR
Q9SE35	36	LSTEDSK SAQLIS
Q9SE35	110	LRILSSK KSGSY Y
Q9SE35	111	RILSSK KSGSY S
Q9SE35	121	YYSYGT KSGSY S
Q9SE35	122	YSYGT K KSGSY S
Q9SE35	132	YSGYST KKSASRR
Q9SE35	133	SGYST KKSASRR I
Q9SE35	177	SGSKGSK RRILSS
Q9SE35	203	RRNLSSK KSGSY S
Q9SE35	204	RNLSSK KSGSY S
O18495	51	AAWQIGK HALGDM
O18495	67	EFQDFMK EVEQAR
